10

A Systematic and Thorough Search for Domains of the Scavenger Receptor Cysteine-Rich Group B Family in the Human Genome

Alexandre M. Carmo[1] and Vattipally B. Sreenu[2]
[1]*IBMC – Instituto de Biologia Molecular e Celular and ICBAS – Instituto de Ciências Biomédicas de Abel Salazar, Universidade do Porto*
[2]*The Weatherall Institute of Molecular Medicine, University of Oxford*
[1]*Portugal*
[2]*UK*

1. Introduction

The biological function of proteins is largely determined by their individual component domains, which are segments within the protein sequence that are self-contained and spatially arranged. These can be catalytic or structural, and define a number of different features of proteins such as their enzymatic activity, interactions with other proteins, sugars or lipids, and determine the cellular localization of the proteins that contain them. A number of intracellular three-dimensionally-arranged domains, such as Src-homology (SH) or Pleckstrin-homology (PH) domains, define the nature of protein interactions with other components of the cell, and enable them to interact with their substrates or binding partners. The specificity of interactions that is given by the domain is unique to its protein. Similarly, the extracellular part of most membrane-bound or secreted proteins of eukaryotic cells is also organized in semi-autonomously-arranged blocks that potentially confer multiple diverse functions to a particular protein. These domains have been classified and grouped into protein superfamilies depending on the similarity they have with domains of prototypical proteins, for example immunoglobulin, fibronectin or C-type lectin domains. Members of these groups are believed to be homologous and to have arisen by divergent evolution from a common ancestor. Many membrane-bound or extracellular proteins are comprised of several domains of the same type, but it is not uncommon to find mosaic proteins containing domains from different superfamilies.

The scavenger receptor cysteine-rich (SRCR) superfamily comprises a group of proteins that contain one or multiple domains structurally similar to the membrane distal domain of the type I scavenger receptor expressed by human macrophages (Freeman et al., 1990). Proteins classified as belonging to this superfamily may contain other types of domains additionally to the dominant SRCR modules, such as EGF, CUB, LCCL, or other domains. In mammals, SRCR proteins are typically expressed in cells of the immune system (Resnick et al., 1994), although some members can be also expressed in non-immune cells and organs, including liver, kidney, placenta, stomach, brain and heart (Sarrias et al., 2004). Group A domain-containing SRCR proteins are present in phyla from the most primitive metazoan to

vertebrates, whereas group B domain containing SRCR proteins are only found in vertebrates. Intriguingly, although SRCR proteins can include other domains, no proteins have been reported to contain group A and B domains simultaneously.
In mammalian species, SRCR group B orthologs are usually very well conserved and regarding some of the proteins, a high level of conservation is extended to birds and fish. However, in some cases a human SRCR protein apparently has no corresponding ortholog in some mammals, and conversely, there are examples of SRCR group B proteins that are well characterized in a few mammalian species, that have not been described in humans. By analyzing the human genome, we can now identify all the remaining, still undescribed genes encoding SRCR group B domains, which will allow us to perform phylogenetic analysis of the complete set of group B domains. By comprehensive and systematic whole genome analysis we have found two new putative transcriptional units containing clusters of potential SRCR domains, and additionally a further putative gene that contains a single domain. After our thorough search, we are now confident that all proteins containing group B SRCR domains in the human genome have been identified.

2. The scavenger receptor cysteine-rich group B family

2.1 Biological function of SRCR group B proteins

The cell surface antigens CD5 and CD6, which function in T lymphocytes, are probably the most well characterized of the family, each containing three extracellular SRCR domains (Aruffo et al., 1991; Jones et al., 1986). CD5 and CD6 co-associate with each other at the surface of T cells (Castro et al., 2003; Gimferrer et al., 2003), and are involved in the regulation of T cell receptor-mediated activation. The extensive characterization of the interaction of CD6 with its ligand CD166, expressed by antigen presenting cells (Aruffo et al., 1997), and the identification of different binding partners for CD5 (Biancone et al., 1996; Calvo et al., 1999; Pospisil et al., 2000; Van de Velde et al., 1991), had initially suggested that SRCR group B domains participate in intercellular contacts *via* protein-protein interactions. Also, the three SRCR domain-containing soluble protein Spα (Gebe et al., 1997) has been reported to bind to cells of myeloid and lymphoid origin. Also known as AIM (apoptosis inhibitor expressed by macrophages), API6 (apoptosis inhibitor 6) or CD5L (CD5-like molecule), Spα is best known for promoting macrophage survival. Therefore, this sub-group of small SRCR-containing proteins may be described as having a role in cellular communication, differentiation and activation. However, for most of the remaining members of the family no such clear function has been established. In particular the lack of cellular ligands for most of these proteins raises the possibility that a totally different function for SRCR domains may exist, if indeed SRCR domain proteins share any common function.
In addition to CD5, CD6 and Spα, the group B SRCR family presently contains five other proteins, of which two, CD163 and M160, are membrane bound and expressed by macrophages. CD163 (Law et al., 1993) and M160 (CD163L1) (Gronlund et al., 2000), which were both identified in human monocytes, are considered a subgroup of the SRCR group B molecules. No definitive function has been established for these molecules, although CD163 has been described as binding to, and internalizing, tumor necrosis factor-like weak inducer of apoptosis (TWEAK), thus having a potential role in atherosclerosis (Moreno et al., 2009). Additionally, CD163 has a detoxifying role in iron metabolism, where by binding to hemoglobin-complexed haptoglobin it is able to remove hemoglobin from the plasma

(Graversen et al., 2002; Kristiansen et al., 2001). The remaining three members of the group B SRCR family are secreted glycoproteins of different sizes and structural complexity. DMBT1, which was identified on the basis of its deletion in a medulloblastoma cell line, is the largest member of the family, comprising 14 SRCR domains separated by SRCR-interspacing domains (Mollenhauer et al., 1997). Apart from being secreted, DMBT1 is also found in association with the plasma membrane of macrophages, although it is not clear whether there is a specific receptor or the poorly characterized DMBT1 gene may encode a transmembrane sequence. Once in the membrane, DMBT1 is a ligand for Surfactant protein D (SP-D), a C-type lectin that binds to exposed carbohydrates (Holmskov et al., 1999). The SRCR soluble proteins S4D-SRCRB and SSc5D have four and five group B domains, respectively, and little is known of their functional or binding properties (Gonçalves et al., 2009; Padilla et al., 2002).

However, it has been recently suggested that Spα (Sarrias et al., 2005), DMBT1 (Bikker et al., 2002), CD163 (Fabriek et al., 2009), CD5 (Vera et al., 2009) and CD6 (Sarrias et al., 2007) are capable of detecting microbe-associated molecular patterns, and could bind and clear bacteria or fungi, reaffirming a scavenger-like role for this group of molecules. These developments notwithstanding, SRCR superfamily proteins may prove to have very diverse functions, to the extent that the structural properties of the highly conserved SRCR domains may be the only unifying feature of the family.

2.2 Structure and organization of SRCR domains

Typically, the 100-110 amino acid-long SRCR domains possess a characteristic pattern of cysteine residues that establish intra-domain disulfide bridges and contribute to the overall architecture of the compact domain. The number of cysteine residues and their distribution, together with the organization of the genomic sequence encoding each domain, divide the SRCR family into two groups, A and B. Group A domains are encoded by split exons, and typically have six cysteine residues establishing three disulfide bonds. Group B domains, on the other hand, are encoded by a single exon and have eight cysteine residues, whose distribution is remarkably conserved in nearly all known domains (Fig. 1).

So far, eight human SRCR group B proteins have been described, Spα, CD5, CD6, S4D, SSc5D, CD163, M160 and DMBT1 that contain three to fourteen SRCR domains. Their encoding genes are dispersed throughout the genome, however a few highly similar pairs such as CD5-CD6 and CD163-M160 are located on the same chromosome. The identity between individual domains of different SRCR group B proteins varies from 20 to 80%, and phylogenetic analysis suggests that they have evolved by sequential intragene duplication, although there are examples that suggest they may have evolved in some cases by inter-protein domain shuffling. Only four SRCR domains have been characterized by X-ray crystallography, and of these three are group A SRCR domains, those of hepsin, a cell surface serine protease involved in cell growth and maintenance of cellular morphology (Somoza et al., 2003), M2bp, a tumor associated antigen and matrix protein (Hohenester et al., 1999), and MARCO, a trimeric SRCR group A protein expressed by macrophages and dendritic cells that recognizes polyanionic particles and pathogens (Ojala et al., 2007). The crystal structure of the membrane proximal domain of CD5 (Rodamilans et al., 2007), together with an NMR solution structure of domain 1 of CD5 (Garza-Garcia et al., 2008) constitute the only sources of structural information of SRCR group B domains. Comparing the structures, it is however apparent that the 3D assembly of the different domains in the two groups is overall conserved, all displaying a very similar fold.

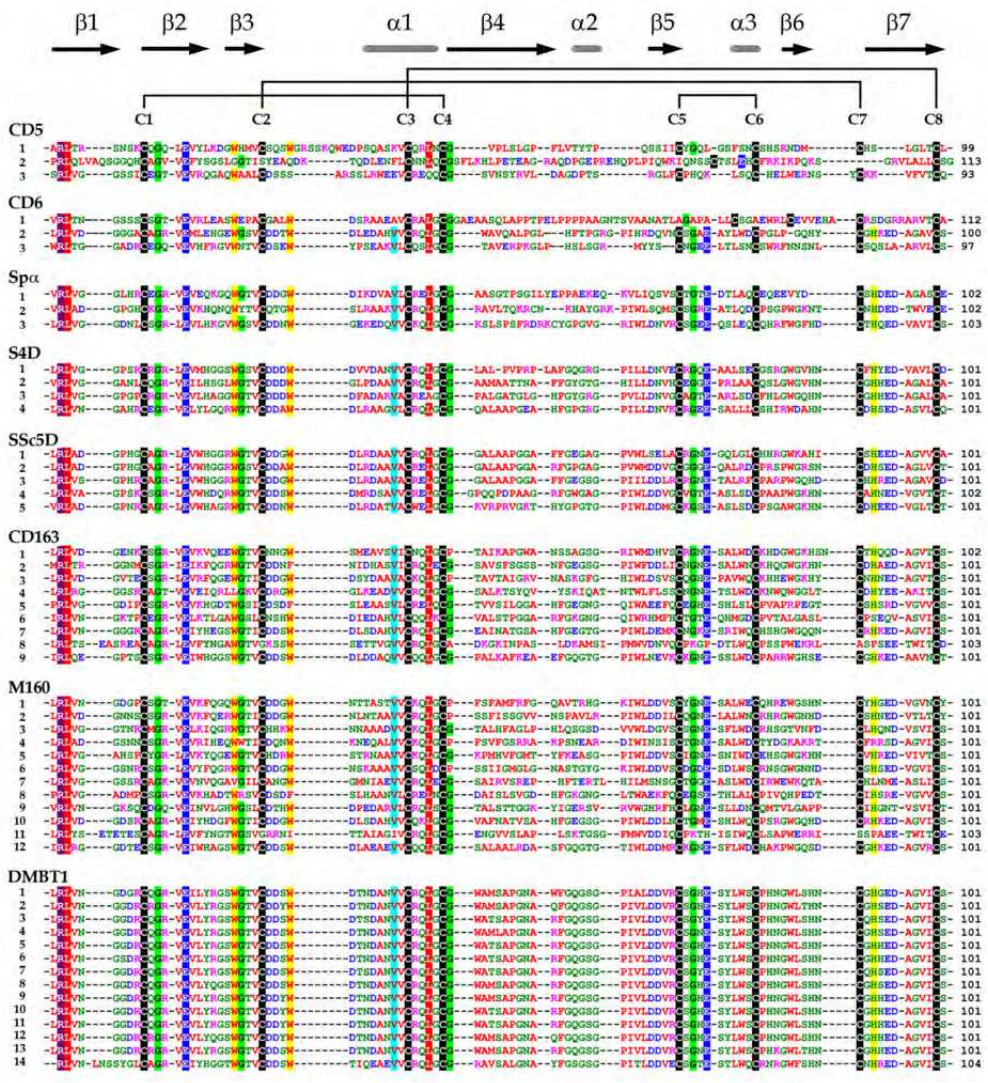


Fig. 1. Sequence alignment of domains from group B SRCR superfamily members.

SRCR domains are typically sequences of 100-110 amino acids in length compacted into a heart-shaped fold, where a six/seven-stranded β-sheet cradles an α-helix. Strands β1, β3 and β4, together with β7, form a curved sheet that wraps around the core α1 helix. From β4 onwards, the structures start to diverge. It is the sequence of amino acids between the beginning of the domain and the β4 strand that is best conserved between group A and group B domains, and that roughly corresponds in the group A proteins to the first of two exons that encode a full SRCR A domain, and in group B proteins to the first 50 amino acids of the domain.

2.3 Homology between SRCR domains

The level of amino acid identity among human SRCR group B domains from different molecules varies from 20% to 80%, but within the same molecule this level can be higher and even be identical in some domains (e.g. domains 3 and 7, and 10 and 11 of DMBT1). Similarly, some molecules are remarkably conserved between species, especially among mammals, although it appears that some level of conservation can be extended to birds, fish and amphibians in a few specific cases. There are good indications for there being orthologs of CD6 in the genomes of *T. guttata* and *D. rerio*, and some other examples. Nevertheless, the structure of SRCR group B-containing molecules is best preserved in mammalian species. The strong homology of SSc5D domains dates back to the divergence of egg- and non-egg-laying mammals, while CD163 has clearly conserved orthologs in all mammals, including non-placental species (Table 1).

	D1	D2	D3	D4	D5	D6	D7	D8	D9
P. troglodytes	99	99	98	97	97	99	99	99	100
C. jacchus	94	87	88	87	95	93	98	89	99
O. cuniculus	90	82	84	81	85	84	73	47	63
R. norvegicus	78	75	82	75	83	81	84	70	79
M. musculus	78	76	77	75	83	76	81	70	79
C. familiaris	88	87	88	90	93	89	92	81	86
E. caballus	85	84	89	87	94	87	92	87	87
B. taurus	82	86	90	87	91	85	89	77	85
M. domestica	51	58	51	45	58	51	76	60	70
O. anatinus	36	28	36	ND	40	57	64	40	70

Table 1. Homology between human and other mammalian CD163 domains. Numbers represent percentage of identity between each domain, compared to the human sequence.

The significantly conserved homology of some SRCR orthologs is suggestive of profound functional constraints acting on these proteins. On the other hand, it appears that not all human SRCR B group molecules have described orthologs in all mammalian species, and conversely, that there are some SRCR proteins described in different animals that have not been reported in man.

Noticeably, bovine WC1 (Wijngaard et al., 1992) does not have a human counterpart, nor do the mouse SCART molecules (Kisielow et al., 2008). Similarly, the human macrophage specific receptor M160 is not found in all mammalian species, while the closely related molecule CD163, also specific to the monocytic/macrophage lineage, is clearly present in all genomes that we have examined. We have compared the similarity between individual domains of known and characterized members of the SRCR B group, and the corresponding domains in the bovine proteins (Table 2). While proteins such as S4D, SSc5D and CD163 show high levels of identity between human and bovine sequences, others like CD5 and Spα are more distantly related. M160 does not have a straightforward ortholog in cattle, so the bovine sequence used was of the related molecule M160-like, that is related in turn to the SCART 1 and 2 molecules present in the mouse.

	CD5	CD6	Spα	S4D	SSc5D	CD163	M160	DMBT1
D1	48	52	ND	99	96	82	ND	ND
D2	48	83	62	95	97	86	ND	ND
D3	66	81	72	97	95	90	32	ND
D4				96	ND	87	ND	86
D5					89	91	28	ND
D6						85	38	86
D7						89	38	ND
D8						77	ND	86
D9						85	32	85
D10							55	88
D11							47	84
D12							69	82
D13								85
D14								83

Table 2. Similarity between human and bovine corresponding SRCR domains. Percentage identity between each domain is indicated.

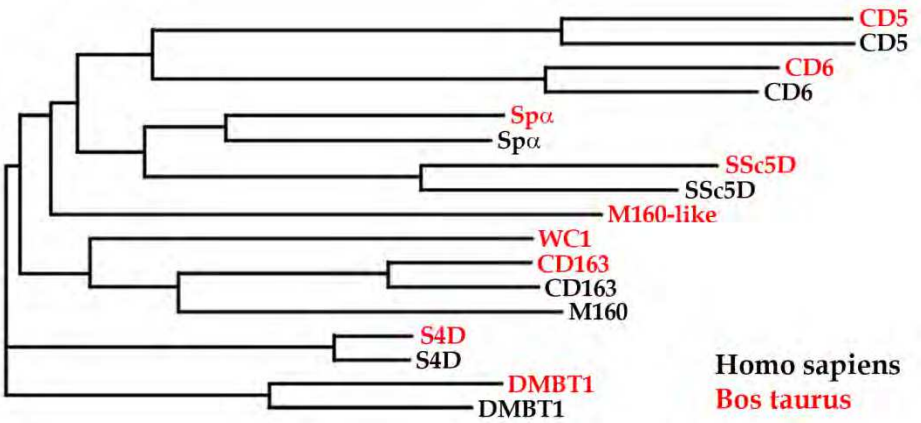


Fig. 2. Relationships between human and bovine SRCR molecules.

There are three groups of genes in the CD163 family: CD163 itself (CD163a), present in all mammals; M160 (CD163b), so far only found in the genomes of primates and in horses; and

SCART (CD163c), of which there are two genes found in the mouse, and the bovine gene M160-like (Herzig et al., 2010). These sets of genes are related to WC1 genes expressed in cattle, sheep and swine. To obtain a better idea of the relationship between these families of genes, we aligned the full sequences of known human and bovine proteins containing SRCR group B domains using ClustalW and drew the corresponding phylogram (Fig. 2). As can be seen, there are no direct links between human M160, bovine M160L and bovine WC1, raising the possibility that either some genes were lost during mammalian evolution, or that the complete characterization and annotation of the genomes has still not been fully achieved. Clearly, either hypothesis does not exclude the other.

3. A systematic and thorough search for SRCR domains in the genome

Our hope is that the evolution and function of SRCR domains would emerge when all members of this protein family have been identified. The advent of the human genome sequence has allowed us to screen, using bioinformatics-based approaches, for new SRCR proteins still not described or characterized. We decided to focus on group B molecules, given that proteins of this type are more conserved, restricted in number, and their specialized function, in this case immune-related, seems better defined. We performed searches for new members of the SRCR-SF in the completed human genome sequence by interrogating the genome using TBLASTN 2.2.20+ (Altschul et al., 1997; http://www.ncbi.nlm.nih.gov/BLAST). Initially, we screened for new sequences exhibiting similarity with any or all of the SRCR domains comprising the then known SRCR superfamily proteins (Gonçalves et al., 2009). We expected that, for a given TBLASTN run, *bona fide* new SRCR domains would have smaller E values than the best matches of the search sequence with Group A SRCR domains. According to this criterion, the search identified the sequences encoding domains within already known and characterized proteins *i.e.* CD5, CD6, Spα, S4D, CD163, M160 and DMBT1. Additionally, we identified a cluster of five new SRCR domains, which we further investigated and that later resulted in the cloning and characterization of SSc5D, a molecule secreted by macrophages and that comprises five SRCR group B domains (Gonçalves et al., 2009).

A *caveat* in our methodology was that not all group B domains were identified using this strategy. The most divergent domains, namely those of CD5, were not retrieved in all searches, and in particular CD5-d1 was rarely identified as having a clear homology with any other group B domain. Sequence alignment of all group B domains (Fig. 1) highlights the striking differences of CD5 sequences, and also to some extent of the CD6 domains, when compared with other sequences that are remarkably similar to each other. In order to perform a more rigorous search for all putative SRCR group B domains, we conducted a comprehensive systematic search using PSI-BLAST (Altschul et al., 1997) to find distant homologs. All known SRCR domains were used as queries to search iteratively against human non-redundant database with the target sequence length set to 250. BLOSUM62 amino acid substitution matrix with gap open penalty 11, and extension penalty 1, was used. Sequence masking was disabled and the PSI-BLAST threshold was set to 0.005. While searching, each PSI-BLAST query was iterated including new hits from the previous search until it converged, i.e. no new hits were found in subsequent searches. After each search iteration, results were checked for new SRCR proteins. This meticulous and robust method picked up all known SRCR domain-containing proteins along with novel proteins (Table 3).

Protein	Number of SRCR Domains	Chromosome
Spα	3	3
S4D	4	7
DMBT1	17*	10
CD5	3	11
CD6	3	11
CD163	9	12
M160	12	12
SSc5D	5	19
8D	8	10
D11#	11	10
HHIPL1	1	14

Table 3. List of SRCR-containing proteins. * - DMBT1 has been described as containing 14 SRCR domains; # - annotated as a pseudogene; in red denotes new SRCR domains from uncharacterized proteins.

From this genome-wide search we obtained a total of 76 SRCR group B domains distributed in 11 genes, each putatively encoding a varying number of SRCR domains. The eleven genes are spread across the genome on seven different chromosomes: chromosome 10 contains three SRCR group B-encoding genes, chromosomes 11 and 12 contain two each, and chromosomes 1, 7, 14 and 19 each contain a copy of just one SRCR-encoding gene. Among these genes and in addition to known domains from characterized genes, our search has uncovered 23 new putative group B domains, three of which represent previously unreported domains localized within the DMBT1 gene. The DMBT1 gene thus putatively encodes a maximum of 17 SRCR domains. Some controversy has existed on the number of SRCR domains within the DMBT1 molecule. Like other SRCR-containing proteins (Castro et al., 2007; Padilla et al., 2002), DMBT1 can be expressed as different isoforms arising by alternative splicing, which include or exclude individual SRCR domains (Mollenhauer et al., 1999). It is possible that the new DMBT1 domains have not been previously reported because they are not expressed in the tissues or cells investigated, however it is also plausible that the exons coding for these domains have been silenced during evolution and are now non-functional.

The remaining new domains belong to 3 new putative genes, one 8 domain-encoding gene (8D), one gene, annotated as a DMBT1-like pseudo-gene, that encodes 11 fragments of SRCR domains of variable lengths (D11), and a gene encoding a putative Hedgehog interacting protein-like 1 molecule (HHIP-like 1), which contains a single SRCR domain.

In order to analyze the sequence conservation and diversity of SRCR domains, we aligned all individual 76 SRCR group B domains using ClustalW2 (Thompson et al., 1994) with the default substitution matrix (Gonnet series) and gap opening and extension penalties of 10 and 0.2 respectively. Due to the sequence diversity in SRCR domains, several insertions and deletions were found in the multiple sequence alignment (Fig. 3).

To locate sequence patterns as well as conserved amino acids, the multiple sequence alignment was used to create a WebLogo (Crooks et al., 2004; http://weblogo.berkeley.edu).

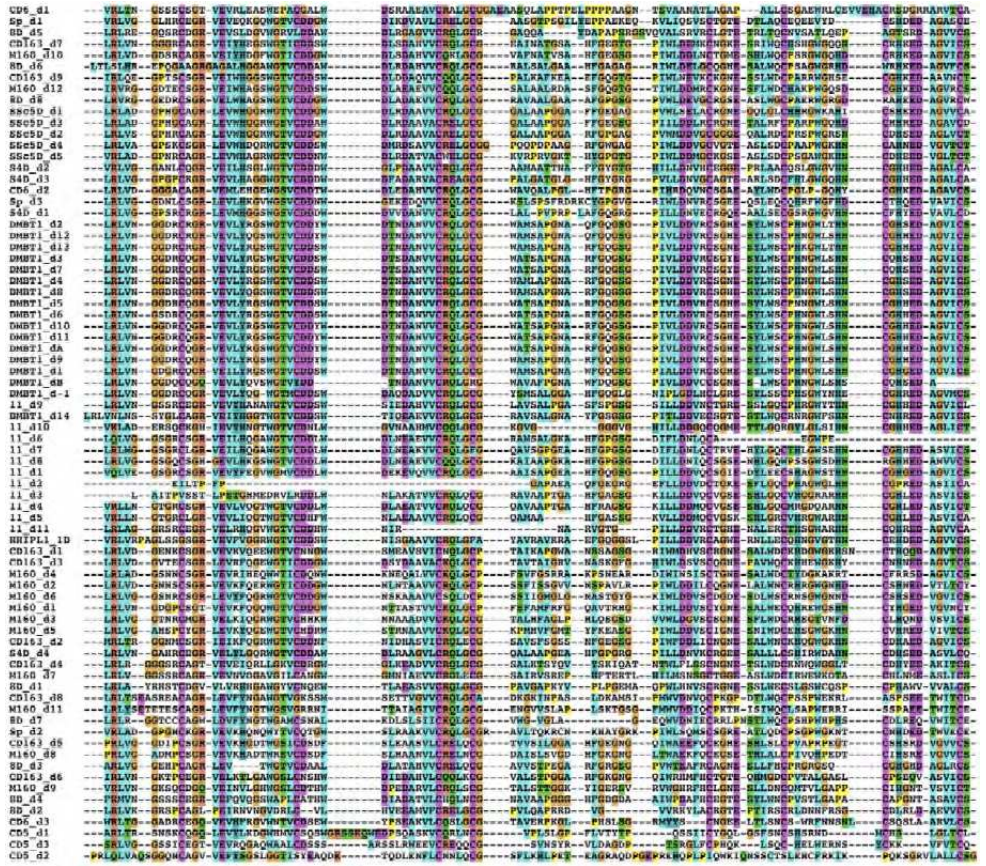


Fig. 3. Multiple sequence alignment of all SRCR domains.

The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino acid at that position. It is apparent from the WebLogo that, although sequences vary substantially between SRCR domains, all cysteine residues (colored in red) are conserved across the family (Fig. 4).

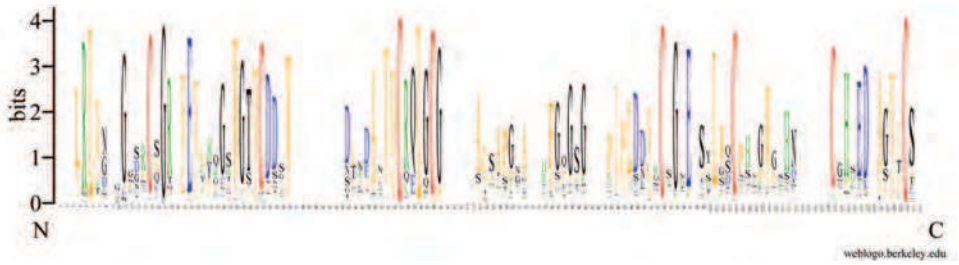


Fig. 4. Multiple sequence alignment of SRCR domains in WebLogo format.

In order to estimate evolutionary relationships among SRCR domains, we performed a phylogenetic analysis utilizing the maximum likelihood phylogenetic reconstruction method (proml) available in the phylip phylogenetic package (Felsenstein, 1989). Gaps were trimmed from the SRCR multiple sequence alignment prior to tree building and the Jones-Taylor-Thornton probability model employed with constant rate of change among sites. The reliability of internal branches was subsequently evaluated using 100 bootstrap samplings. SRCR domains exhibit very complex evolutionary relationships (Fig. 5). In the reconstructed phylogenetic tree, intra-protein domain clustering as well as inter-protein domain clusters were observed. Intra-domain clustering, as in the case of DMBT1, strongly suggests the evolution of these domains via sequential intragenic duplication. At the same time it is difficult to understand the inter-protein domain similarities. CD5, CD6, Spα, M160 and CD163 exhibit more diverse relationships. Among them, SRCR domains show greater inter than intra protein similarities. Given their low sequence similarities, it is uncertain whether the domains evolved through gene duplication and accumulated mutations have reduced sequence similarity, or if it is through a convergent evolution mechanism subsequent to domain shuffling. The similarities of domain pairs M160_d4-CD163_d1, M160_d7-CD163_d4 M160_d8-CD163_d5, M160_d9-CD163_d6, M160_d10-CD163_d7, M160_d11-CD163_d8, M160_d12–CD163_d9, CD5_d1-CD6_d3 and CD5_d2-S4D_d4 are strongly suggestive of inter-protein domain shuffling.

4. Concluding remarks - the completion of the SRCR group B family

In contrast to the complexity and variety of large protein families such as the G protein coupled receptor (GPCR) superfamily, which has nearly 800 genes in the human genome, corresponding to roughly 4% of the full protein-encoding genome, group B of the scavenger receptor cysteine-rich superfamily appears to be much more limited. So far it includes only 8 members in the entire human genome, although there are additionally 3 proteins described in other mammalians; SCART1 and SCART2 initially found in mice, and the 11 SRCR domain-containing protein WC1 expressed in cattle, sheep and swine. Also, the function of mammalian SRCR proteins seems to be restricted to the immune system, although the exact nature or biological role of the family is still to be fully determined.

In this study we set out to identify the remaining members of the SRCR group B family in order to obtain a clear understanding of the biological significance of this important group of proteins and to clarify some as yet unresolved questions regarding their evolution in mammalian species. We searched the human genome for the presence of SRCR-encoding genes using as probes the amino acid sequence of all reported human SRCR domains. Interestingly, one of the new members we have identified, HHIP-like 1, contains a single SRCR domain, which is unknown in the family. Moreover, the amino acid sequence corresponding to the SRCR domain constitutes only a small fraction of the total of the putative protein (13%). This is in contrast with most other members, whose amino acid content corresponding to SRCR domains relative to the whole of the protein is significantly higher, varying between 32% (SSc5D) and 92% (Spα). HHIP-like 1 is related to Hedgehog interacting protein, a regulatory component of the Hedgehog signaling pathway (Chuang and McMahon, 1999). However, unlike HHIP-like 1, HHIP does not contain an SRCR domain.

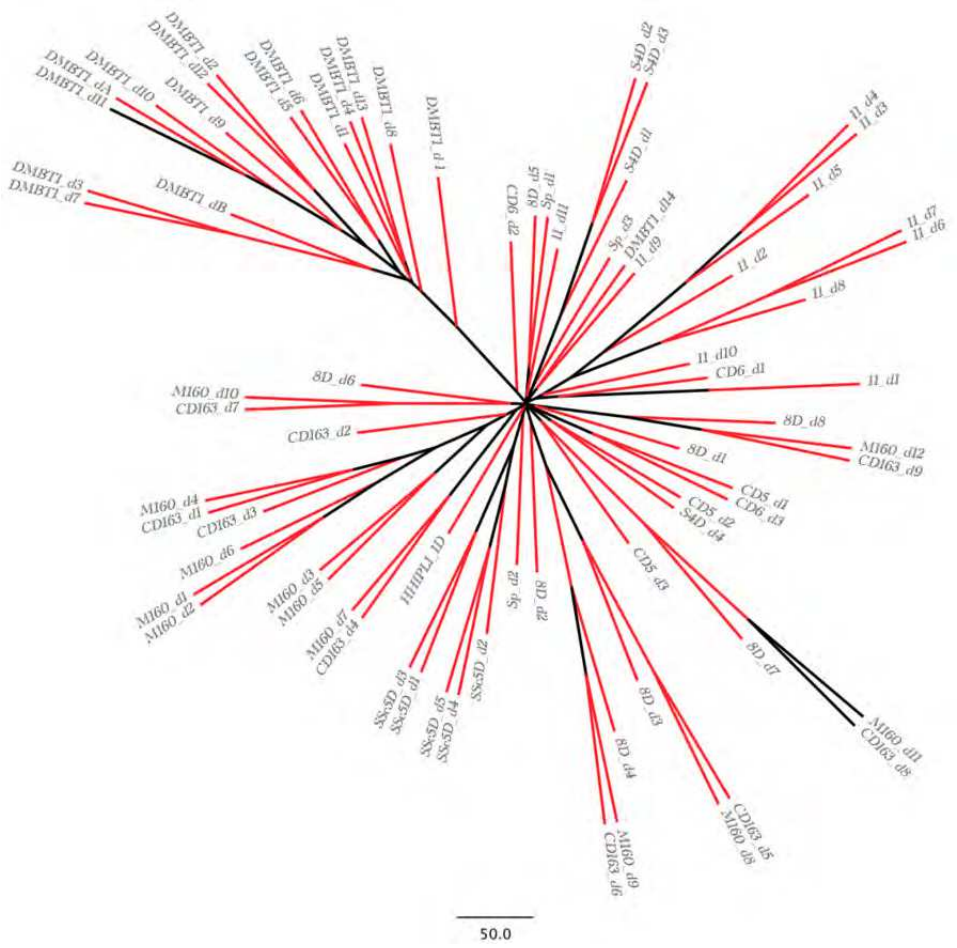


Fig. 5. Maximum likelihood phylogenetic estimation of human SRCR domains. Internal branch reliability was assessed using bootstrapping method (100 bootstrap replicates). Branches observed in more than 75 of 100 bootstrapped re-sampling are shown in red.

The second cluster of SRCR domains we have uncovered is located on chromosome 10 and includes 8 such domains, thus we provisionally termed it 8D. Using the predicted exon-derived protein sequence, we BLAST-searched other mammalian genomes and the proteins that we retrieved which were most similar to human 8D were mouse SCART1, bovine M160L, and mouse SCART2, whose ClustalW alignment scores were 70, 66 and 57, respectively (Fig. 6). Human 8D and mouse SCART1 have 64% identity for the entire sequence, while some individual SRCR domains share identities of close to, and even above 80%. We thus believe that 8D is the human ortholog of mouse SCART1. It remains to be seen whether human 8D can be expressed and produce a mature and functional protein, although we have detected several 8D transcripts of different sizes (C Gonçalves and A Carmo, unpublished).

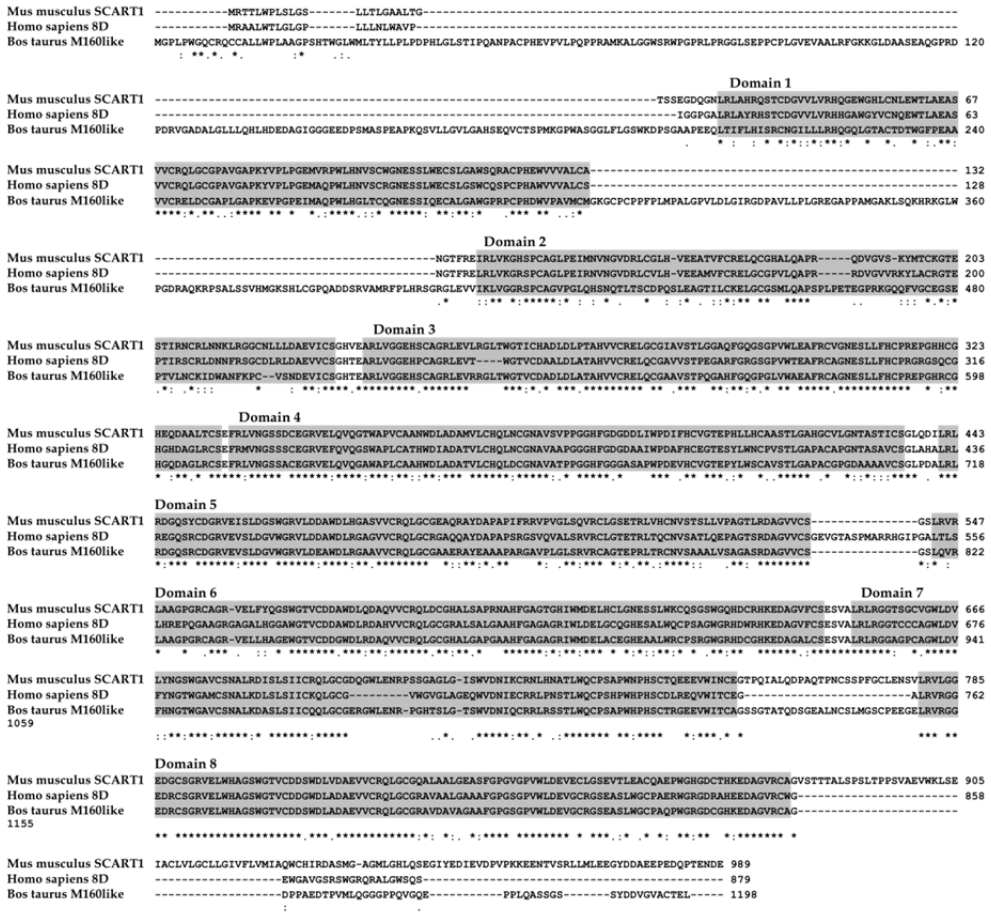


Fig. 6. Sequence alignment of mouse SCART1, human 8D and bovine M160-like protein.

The last new set of domains we have identified is located in a gene also on chromosome 10, but has been annotated as a non-coding pseudogene. Analysis of its putative sequence derived from the exon-like sequences in fact reveal that some stretches of several of the SRCR domains are missing, adding to a number of frameshifts and premature stop codons. Curiously, 11 SRCR-like domains can be identified, exactly the same number as the typical bovine WC1 protein. Comparison between the two sequences has failed however to definitely determine whether these two genes have the same evolutionary origin, as individually identifiable SRCR or SRCR-type domains seem to have already drifted apart significantly.

With the recognition of the three new genes, albeit none of them proven to be functional as yet together with the detection of three new putative SRCR-encoding sequences present in the DMBT1 gene, we are confident that we have completed the identification of the full set of scavenger receptor cysteine-rich group B domains in the human genome.

5. Acknowledgements

We thank Dr. Simon Lee for reviewing this manuscript. The work in Alexandre Carmo's laboratory is funded by FEDER through the Programa Operacional Factores de Competitividade – COMPETE, and by FCT – Fundação para a Ciência e a Tecnologia. Vattipally Sreenu is funded by the University of Oxford.

6. References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res *25*, 3389-3402.

Aruffo, A., Bowen, M.A., Patel, D.D., Haynes, B.F., Starling, G.C., Gebe, J.A., and Bajorath, J. (1997). CD6-ligand interactions: a paradigm for SRCR domain function? Immunol Today *18*, 498-504.

Aruffo, A., Melnick, M.B., Linsley, P.S., and Seed, B. (1991). The lymphocyte glycoprotein CD6 contains a repeated domain structure characteristic of a new family of cell surface and secreted proteins. J Exp Med *174*, 949-952.

Biancone, L., Bowen, M.A., Lim, A., Aruffo, A., Andres, G., and Stamenkovic, I. (1996). Identification of a novel inducible cell-surface ligand of CD5 on activated lymphocytes. J Exp Med *184*, 811-819.

Bikker, F.J., Ligtenberg, A.J., Nazmi, K., Veerman, E.C., van't Hof, W., Bolscher, J.G., Poustka, A., Nieuw Amerongen, A.V., and Mollenhauer, J. (2002). Identification of the bacteria-binding peptide domain on salivary agglutinin (gp-340/DMBT1), a member of the scavenger receptor cysteine-rich superfamily. J Biol Chem *277*, 32109-32115.

Calvo, J., Places, L., Padilla, O., Vilà, J.M., Vives, J., Bowen, M.A., and Lozano, F. (1999). Interaction of recombinant and natural soluble CD5 forms with an alternative cell surface ligand. Eur J Immunol *29*, 2119-2129.

Castro, M.A.A., Nunes, R.J., Oliveira, M.I., Tavares, P.A., Simões, C., Parnes, J.R., Moreira, A., and Carmo, A.M. (2003). OX52 is the rat homologue of CD6: evidence for an effector function in the regulation of CD5 phosphorylation. Journal of Leukocyte Biology *73*, 183-190.

Castro, M.A.A., Oliveira, M.I., Nunes, R.J., Fabre, S., Barbosa, R., Peixoto, A., Brown, M.H., Parnes, J.R., Bismuth, G., Moreira, A., *et al*. (2007). Extracellular Isoforms of CD6 generated by alternative splicing regulate targeting of CD6 to the immunological synapse. Journal of Immunology *178*, 4351-4361.

Chuang, P.T., and McMahon, A.P. (1999). Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. Nature *397*, 617-621.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Fabriek, B.O., van Bruggen, R., Deng, D.M., Ligtenberg, A.J., Nazmi, K., Schornagel, K., Vloet, R.P., Dijkstra, C.D., and van den Berg, T.K. (2009). The macrophage scavenger receptor CD163 functions as an innate immune sensor for bacteria. Blood *113*, 887-892.

Felsenstein, J. (1989). Mathematics vs. Evolution: Mathematical Evolutionary Theory. Science *246*, 941-942.

Freeman, M., Ashkenas, J., Rees, D.J., Kingsley, D.M., Copeland, N.G., Jenkins, N.A., and Krieger, M. (1990). An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II murine macrophage scavenger receptors. Proc Natl Acad Sci USA *87*, 8810-8814.

Garza-Garcia, A., Esposito, D., Rieping, W., Harris, R., Briggs, C., Brown, M.H., and Driscoll, P.C. (2008). Three-dimensional solution structure and conformational plasticity of the N-terminal scavenger receptor cysteine-rich domain of human CD5. J Mol Biol *378*, 129-144.

Gebe, J.A., Kiener, P.A., Ring, H.Z., Li, X., Francke, U., and Aruffo, A. (1997). Molecular cloning, mapping to human chromosome 1 q21-q23, and cell binding characteristics of Spalpha, a new member of the scavenger receptor cysteine-rich (SRCR) family of proteins. J Biol Chem *272*, 6151-6158.

Gimferrer, I., Farnós, M., Calvo, M., Mittelbrunn, M., Enrich, C., Sánchez-Madrid, F., Vives, J., and Lozano, F. (2003). The accessory molecules CD5 and CD6 associate on the membrane of lymphoid T cells. J Biol Chem *278*, 8564-8571.

Gonçalves, C.M., Castro, M.A.A., M., Henriques, T., Oliveira, M.I., Pinheiro, H.C., H., Oliveira, C., Sreenu, V.B., Evans, E.J., Davis, S.J., Moreira, A., *et al.* (2009). Molecular cloning and analysis of SSc5D, a new member of the scavenger receptor cysteine-rich superfamily. Molecular Immunology *46*, 2585-2596.

Graversen, J.H., Madsen, M., and Moestrup, S.K. (2002). CD163: a signal receptor scavenging haptoglobin-hemoglobin complexes from plasma. Int J Biochem Cell Biol *34*, 309-314.

Gronlund, J., Vitved, L., Lausen, M., Skjodt, K., and Holmskov, U. (2000). Cloning of a novel scavenger receptor cysteine-rich type I transmembrane molecule (M160) expressed by human macrophages. J Immunol *165*, 6406-6415.

Herzig, C.T., Waters, R.W., Baldwin, C.L., and Telfer, J.C. (2010). Evolution of the CD163 family and its relationship to the bovine gamma delta T cell co-receptor WC1. BMC Evol Biol *10*, 181.

Hohenester, E., Sasaki, T., and Timpl, R. (1999). Crystal structure of a scavenger receptor cysteine-rich domain sheds light on an ancient superfamily. Nat Struct Biol *6*, 228-232.

Holmskov, U., Mollenhauer, J., Madsen, J., Vitved, L., Gronlund, J., Tornoe, I., Kliem, A., Reid, K.B., Poustka, A., and Skjodt, K. (1999). Cloning of gp-340, a putative opsonin receptor for lung surfactant protein D. Proc Natl Acad Sci USA *96*, 10794-10799.

Jones, N.H., Clabby, M.L., Dialynas, D.P., Huang, H.J., Herzenberg, L.A., and Strominger, J.L. (1986). Isolation of complementary DNA clones encoding the human lymphocyte glycoprotein T1/Leu-1. Nature *323*, 346-349.

Kisielow, J., Kopf, M., and Karjalainen, K. (2008). SCART scavenger receptors identify a novel subset of adult gammadelta T cells. J Immunol *181*, 1710-1716.

Kristiansen, M., Graversen, J.H., Jacobsen, C., Sonne, O., Hoffman, H.J., Law, S.K., and Moestrup, S.K. (2001). Identification of the haemoglobin scavenger receptor. Nature *409*, 198-201.

Law, S.K., Micklem, K.J., Shaw, J.M., Zhang, X.P., Dong, Y., Willis, A.C., and Mason, D.Y. (1993). A new macrophage differentiation antigen which is a member of the scavenger receptor superfamily. Eur J Immunol *23*, 2320-2325.

Mollenhauer, J., Holmskov, U., Wiemann, S., Krebs, I., Herbertz, S., Madsen, J., Kioschis, P., Coy, J.F., and Poustka, A. (1999). The genomic structure of the DMBT1 gene: evidence for a region with susceptibility to genomic instability. Oncogene *18*, 6233-6240.

Mollenhauer, J., Wiemann, S., Scheurlen, W., Korn, B., Hayashi, Y., Wilgenbus, K.K., von Deimling, A., and Poustka, A. (1997). DMBT1, a new member of the SRCR superfamily, on chromosome 10q25.3-26.1 is deleted in malignant brain tumours. Nat Genet *17*, 32-39.

Moreno, J.A., Muñoz-García, B., Martín-Ventura, J.L., Madrigal-Matute, J., Orbe, J., Páramo, J.A., Ortega, L., Egido, J., and Blanco-Colio, L.M. (2009). The CD163-expressing macrophages recognize and internalize TWEAK: potential consequences in atherosclerosis. Atherosclerosis *207*, 103-110.

Ojala, J.R., Pikkarainen, T., Tuuttila, A., Sandalova, T., and Tryggvason, K. (2007). Crystal structure of the cysteine-rich domain of scavenger receptor MARCO reveals the presence of a basic and an acidic cluster that both contribute to ligand recognition. J Biol Chem *282*, 16654-16666.

Padilla, O., Pujana, M.A., López-de la Iglesia, A., Gimferrer, I., Arman, M., Vilà, J.M., Places, L., Vives, J., Estivill, X., and Lozano, F. (2002). Cloning of S4D-SRCRB, a new soluble member of the group B scavenger receptor cysteine-rich family (SRCR-SF) mapping to human chromosome 7q11.23. Immunogenetics *54*, 621-634.

Pospisil, R., Silverman, G.J., Marti, G.E., Aruffo, A., Bowen, M.A., and Mage, R.G. (2000). CD5 is a potential selecting ligand for B-cell surface immunoglobulin: a possible role in maintenance and selective expansion of normal and malignant B cells. Leuk Lymphoma *36*, 353-365.

Resnick, D., Pearson, A., and Krieger, M. (1994). The SRCR superfamily: a family reminiscent of the Ig superfamily. Trends Biochem Sci *19*, 5-8.

Rodamilans, B., Muñoz, I.G., Bragado-Nilsson, E., Sarrias, M.R., Padilla, O., Blanco, F.J., Lozano, F., and Montoya, G. (2007). Crystal structure of the third extracellular domain of CD5 reveals the fold of a group B scavenger cysteine-rich receptor domain. J Biol Chem *282*, 12669-12677.

Sarrias, M.R., Roselló, S., Sánchez-Barbero, F., Sierra, J.M., Vila, J., Yélamos, J., Vives, J., Casals, C., and Lozano, F. (2005). A role for human Sp alpha as a pattern recognition receptor. J Biol Chem *280*, 35391-35398.

Sarrias, M.R., Farnós, M., Mota, R., Sánchez-Barbero, F., Ibáñez, A., Gimferrer, I., Vera, J., Fenutría, R., Casals, C., Yélamos, J., *et al.* (2007). CD6 binds to pathogen-associated molecular patterns and protects from LPS-induced septic shock. Proc Natl Acad Sci USA *104*, 11724-11729.

Sarrias, M.R., Grønlund, J., Padilla, O., Madsen, J., Holmskov, U., and Lozano, F. (2004). The Scavenger Receptor Cysteine-Rich (SRCR) domain: an ancient and highly conserved protein module of the innate immune system. Crit Rev Immunol *24*, 1-37.

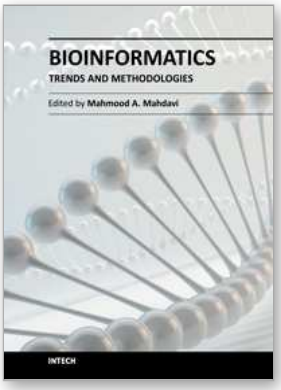
Somoza, J.R., Ho, J.D., Luong, C., Ghate, M., Sprengeler, P.A., Mortara, K., Shrader, W.D., Sperandio, D., Chan, H., McGrath, M.E., *et al.* (2003). The structure of the extracellular region of human hepsin reveals a serine protease domain and a novel scavenger receptor cysteine-rich (SRCR) domain. Structure *11*, 1123-1131.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res *22*, 4673-4680.

Van de Velde, H., von Hoegen, I., Luo, W., Parnes, J.R., and Thielemans, K. (1991). The B-cell surface protein CD72/Lyb-2 is the ligand for CD5. Nature *351*, 662-665.

Vera, J., Fenutría, R., Cañadas, O., Figueras, M., Mota, R., Sarrias, M.R., Williams, D.L., Casals, C., Yelamos, J., and Lozano, F. (2009). The CD5 ectodomain interacts with conserved fungal cell wall components and protects from zymosan-induced septic shock-like syndrome. Proc Natl Acad Sci USA *106*, 1506-1511.

Wijngaard, P.L., Metzelaar, M.J., MacHugh, N.D., Morrison, W.I., and Clevers, H.C. (1992). Molecular characterization of the WC1 antigen expressed specifically on bovine CD4-CD8- gamma delta T lymphocytes. J Immunol *149*, 3273-3277.



Bioinformatics - Trends and Methodologies
Edited by Dr. Mahmood A. Mahdavi

ISBN 978-953-307-282-1
Hard cover, 722 pages
Publisher InTech
Published online 02, November, 2011
Published in print edition November, 2011

Bioinformatics - Trends and Methodologies is a collection of different views on most recent topics and basic concepts in bioinformatics. This book suits young researchers who seek basic fundamentals of bioinformatic skills such as data mining, data integration, sequence analysis and gene expression analysis as well as scientists who are interested in current research in computational biology and bioinformatics including next generation sequencing, transcriptional analysis and drug design. Because of the rapid development of new technologies in molecular biology, new bioinformatic techniques emerge accordingly to keep the pace of in silico development of life science. This book focuses partly on such new techniques and their applications in biomedical science. These techniques maybe useful in identification of some diseases and cellular disorders and narrow down the number of experiments required for medical diagnostic.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821